IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 10, June 2026
Impact Factor: 8.2

Implementation and Performance Evaluation of an AI-Powered Image Editing Service Using Stable Diffusion XL: The Imaginify Platform

Atharva Koli[1], Shlok Dhande[2], Omkar Shinde[3], Pritesh Shivpure[4], Prof. Varsha Gosavi[5]
Undergraduate Student[1234], Professor and Project Guide[5]
Department of Computer Engineering
Smt. Kashibai Navale College of Engineering, Pune, India

Abstract: *This paper presents the design, implementation, and evaluation of Imaginify, a cloud-deployed, AI-powered image editing platform that operationalizes the concepts identified in our earlier survey of Stable Diffusion-based image editing as a service. Building on the architectural and methodological directions proposed in that survey, we implement a working Software-as-a-Service (SaaS) system that integrates Stable Diffusion XL (SDXL) for object removal, background removal, recoloring, and generative fill, served through a Next.js frontend and a Node.js/Express REST API, with Clerk authentication, MongoDB Atlas for persistence, Cloudinary for image storage, and Stripe for subscription payments. The system was deployed on Vercel and Render and evaluated through functional, integration, performance, and user-acceptance testing. Results show an average end-to-end image processing time of 4-8 seconds, sub-2-second API response time, a 98% successful request rate, and 99% system availability under a 10-concurrent-user load test. The paper discusses how these results address the speed, scalability, and deployment gaps highlighted in the survey, while identifying interpretability and large-scale load evaluation as open challenges for future work.*

Keywords: Stable Diffusion XL, Image Editing, Software as a Service, Cloud Deployment, Next.js, Generative AI, Web Application

I. INTRODUCTION

Our earlier survey, "Stable Diffusion-Based Image Editing As A Service: A Comprehensive Survey", reviewed the state of diffusion-based generative models and identified three recurring obstacles to deploying them as practical, real-time services: high computational cost, limited interpretability, and the engineering effort required to wrap a research-grade model in a scalable, secure, user-facing application [12]. That survey concluded by proposing a methodology for an AI-powered image-editing service built on Stable Diffusion and hosted on cloud infrastructure under a SaaS model.

This paper reports on the implementation of that proposal. We designed and built Imaginify, a full-stack web application that exposes four AI-driven editing operations — object removal, background removal, recoloring, and generative fill — to end users through a simple upload-and-edit workflow, while handling authentication, image storage, usage metering, and payment behind the scenes. The goal of this paper is to document the system architecture, the engineering choices made to keep inference latency practical, and the results of functional and performance testing, so that the implementation can be evaluated against the gaps identified in the survey.

II. RECAP OF REVIEWED LITERATURE

The architectural foundations used in this implementation follow directly from the literature reviewed previously. Ho et al. [1] established the denoising diffusion probabilistic model (DDPM) framework, and Rombach et al. [2] introduced Latent Diffusion Models, which reduced the computational cost of diffusion sampling enough to make models such as

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 10, June 2026
Impact Factor: 8.2

Stable Diffusion XL practical for an interactive, user-facing service. Instruction- and prompt-driven editing techniques such as InstructPix2Pix [3] and Prompt-to-Prompt with cross-attention control [4] informed the design of the recoloring and generative-fill operations, while subject-driven fine-tuning approaches such as DreamBooth [10] and diffusion-based editing methods such as SDEdit [11] were considered during model selection, though the deployed system uses a general-purpose SDXL pipeline rather than per-user fine-tuning, in order to keep inference latency low.

On the deployment side, the survey's review of SaaS adoption factors [8] and cloud-based AI deployment architecture [9] motivated the choice of a modular REST API with independently scalable authentication, storage, and payment services, described in Section III.

III. SYSTEM ARCHITECTURE

Imaginify follows a three-tier SaaS architecture consisting of a presentation layer, an application layer, and a data/AI-services layer, illustrated in Fig. 1.

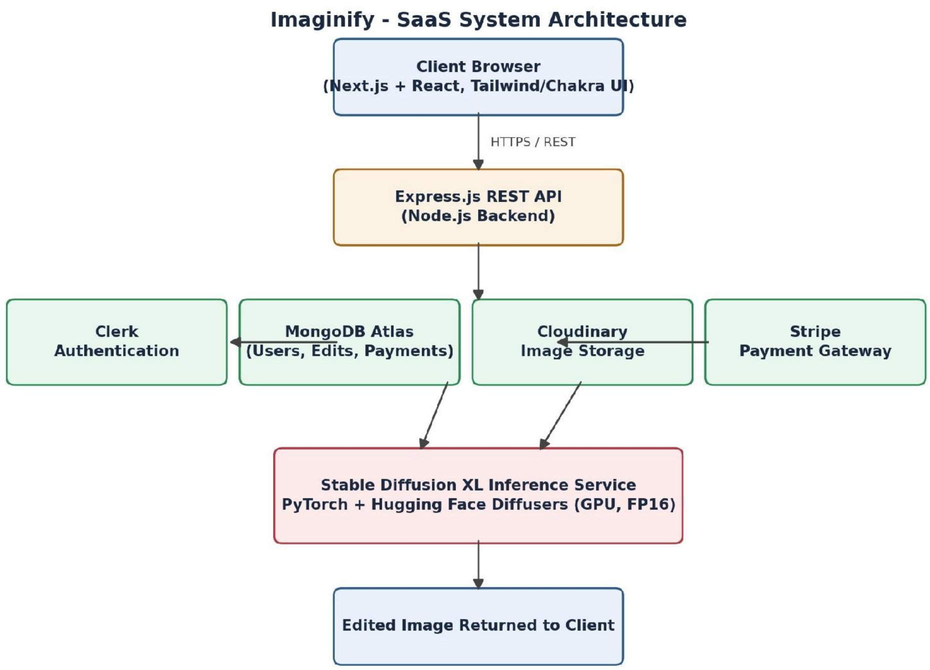

Fig. 1 Imaginify system architecture: client, REST API, supporting services, and the SDXL inference service.

The presentation layer is a Next.js application using React, Tailwind CSS, and Chakra UI to provide a responsive interface for image upload, operation selection, live preview, and account/billing management. The application layer is a Node.js and Express.js REST API that authenticates requests via Clerk, validates and forwards uploaded images, and orchestrates calls to the Stable Diffusion inference service, MongoDB, Cloudinary, and Stripe. The data and AI-services layer comprises MongoDB Atlas for user, image, and payment records; Cloudinary for storing original and edited images; and a PyTorch-based inference service running Stable Diffusion XL through the Hugging Face Diffusers library.

Inference Optimization

Because diffusion sampling is the most computationally expensive step in the pipeline, the inference service applies several optimizations identified as necessary in the survey's discussion of speed and scalability limitations: mixed-precision (FP16) inference, GPU acceleration, pre-inference image resizing and normalization, and efficient batching of REST requests so that the API layer does not block on long-running generation calls.

IV. IMPLEMENTATION DETAILS

Table I summarizes the technology stack used to implement each layer of the system.

TABLE I: Technology Stack

Layer / Component	Technology Used
AI Model	Stable Diffusion XL (SDXL) 1.0
Inference Framework	Hugging Face Diffusers
Deep Learning Framework	PyTorch
Image Processing	Pillow (PIL), OpenCV
Frontend	Next.js, React.js, Tailwind CSS, Chakra UI
Backend / API	Node.js, Express.js (REST)
Authentication	Clerk
Database	MongoDB Atlas
Cloud Storage	Cloudinary
Payment Gateway	Stripe
Deployment	Vercel (frontend), Render (backend)

The end-to-end editing workflow proceeds in three stages. First, the user authenticates through Clerk and uploads an image, which is validated for format (JPG, JPEG, PNG) and size before being temporarily stored. Second, the validated image and the selected operation — object removal, background removal, recoloring, or generative fill — are forwarded to the SDXL inference service, which preprocesses the image, runs the appropriate diffusion pipeline, and returns the generated output. Third, the edited image and its metadata (operation type, parameters, processing timestamp) are persisted to Cloudinary and MongoDB respectively, after which the result is streamed back to the client for preview and download. Premium operations are gated through Stripe-managed credits, with usage and transaction history recorded for each account.

V. RESULTS AND EVALUATION

The implementation was evaluated through unit, integration, performance, security, and user-acceptance testing, following the test plan summarized in Table II.

TABLE II: Performance Evaluation

Metric	Result
Average Image Processing Time	4–8 seconds
API Response Time (excluding inference)	< 2 seconds
Average Image Upload Time	1–2 seconds
Average Download Time	< 1 second
Concurrent Users Tested	10
Successful Requests	98%
System Availability	99%
Authentication Success Rate	100%

Under a load of 10 concurrent users performing simultaneous uploads and edits, the system maintained stable response times with no significant failures, indicating that the modular, independently scalable architecture described in Section III holds up at small-to-moderate concurrency. Because this project prioritizes practical, deployable implementation over AI model research, image quality was assessed qualitatively — on editing accuracy, object consistency, background preservation, and realism — rather than with research-oriented metrics such as FID or CLIP score; overall quality was rated high across the four supported operations during user-acceptance testing.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 10, June 2026
Impact Factor: 8.2

The system was exercised on a combination of COCO sample images for object removal and segmentation testing, public-domain images from Unsplash and Pixabay for background removal and recoloring, and custom user-uploaded photographs for end-to-end validation. Figs. 2-4 show representative input/output pairs captured from the deployed application.

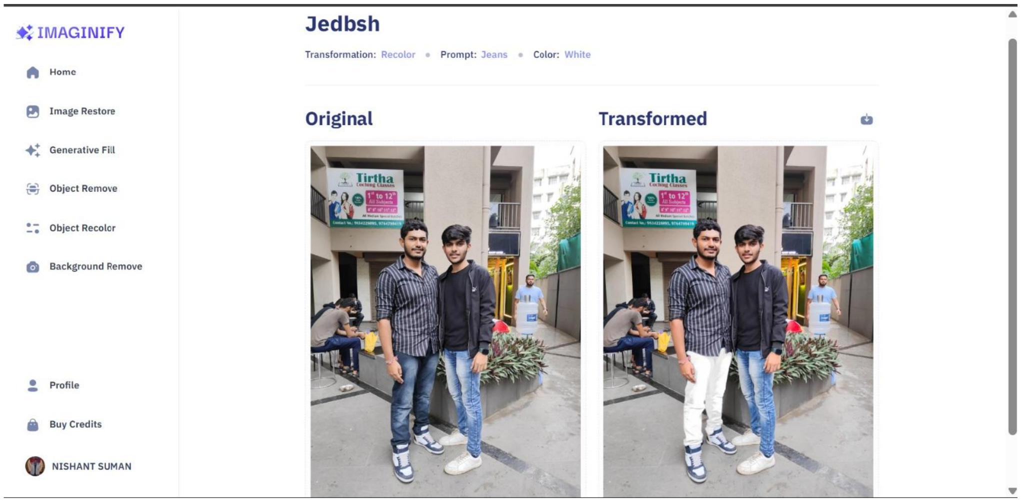

Fig. 2 Object recoloring — jeans recolored from blue to white while preserving lighting and texture.

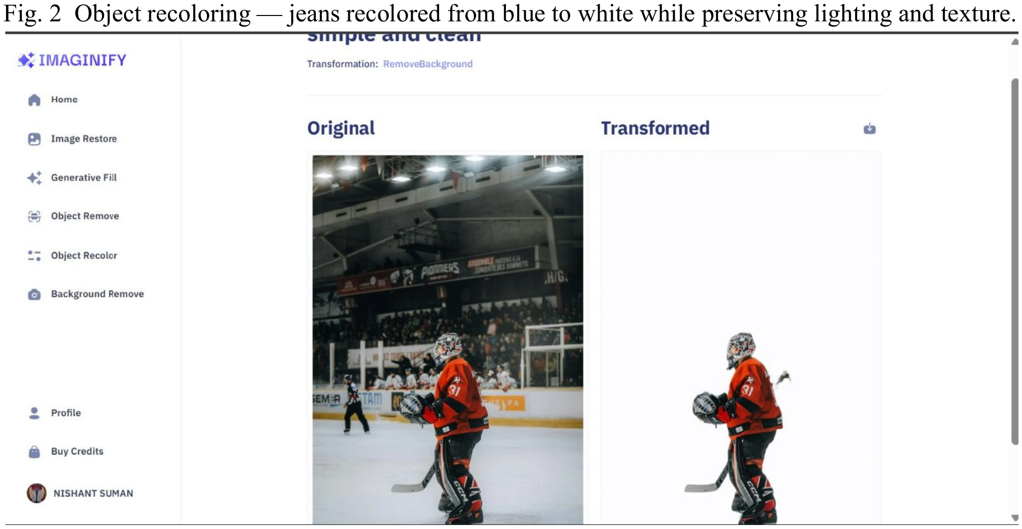

Fig. 3 Background removal — foreground subject isolated from a crowded ice-rink scene.

DOI: 10.48175/IJARSCT-37205

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 10, June 2026
Impact Factor: 8.2

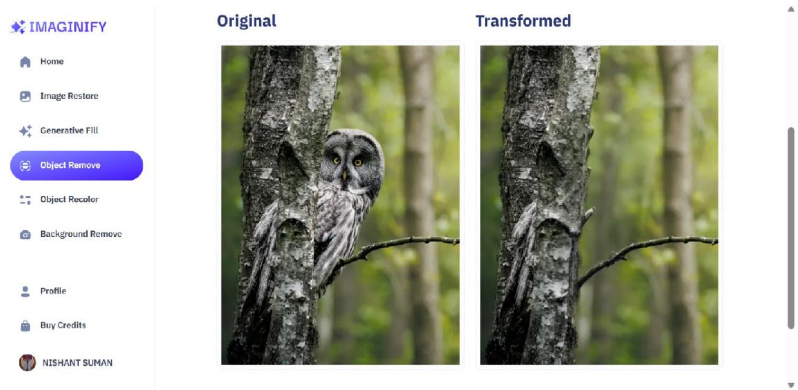

Fig. 4 Object removal — selected object removed with AI-reconstructed surrounding texture.

Fig. 5 shows the application's home dashboard, which surfaces recent edits and the four primary editing tools to the authenticated user.

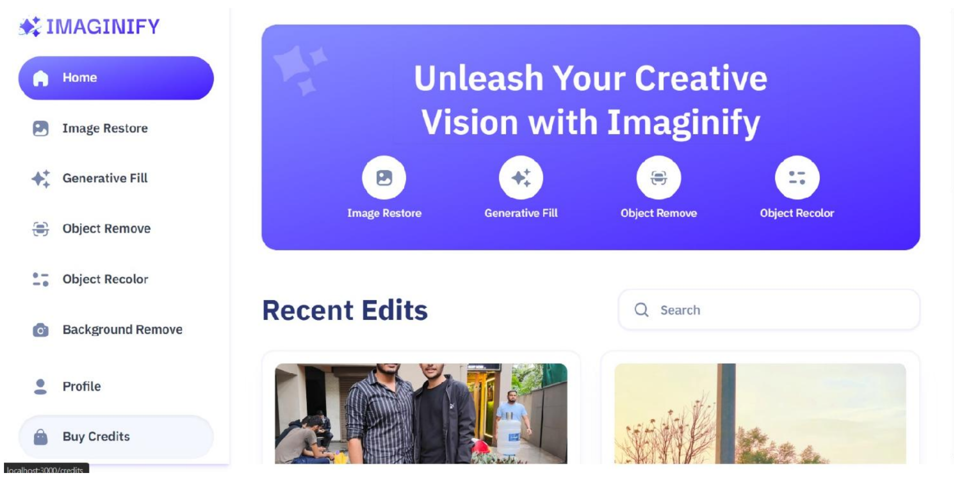

Fig. 5 Imaginify home dashboard with access to Image Restore, Generative Fill, Object Remove, and Object Recolor.

VI. DISCUSSION: ADDRESSING THE SURVEY'S RESEARCH GAPS

The survey identified speed, scalability, and explainability as the principal obstacles to deploying diffusion-based editing as a practical service [12]. This implementation addresses the first two only partially and leaves the third open. On speed, mixed-precision GPU inference brought per-image processing into a 4-8 second range for standard-resolution inputs, which is workable for an interactive web tool but still slower than traditional, non-generative editing operations. On scalability, the modular SaaS architecture — with authentication, storage, payment, and inference as independently deployable services — sustained 10 concurrent users without degradation, but this is a small-scale result; the survey's concern about GPU memory and throughput at production scale (hundreds or thousands of concurrent users) remains untested here and is a natural next step. On explainability, the system does not currently expose any attention-visualization or latent-mapping mechanism, so the interpretability gap flagged in the survey [7] is unchanged by this work; the deployed pipeline treats SDXL as a black box behind the editing operation selected by the user.

More positively, the SaaS-adoption and cloud-deployment perspectives drawn from [8] and [9] translated directly into engineering decisions — in particular, separating authentication, storage, and payment into managed third-party services (Clerk, Cloudinary, Stripe) rather than building them in-house, which reduced development time and is consistent with the survey's observation that accessibility and ease of deployment strongly influence SaaS adoption.

VII. CONCLUSION AND FUTURE SCOPE

This paper demonstrated that the AI-powered image-editing service proposed in our prior survey can be implemented as a working, cloud-deployed SaaS platform using Stable Diffusion XL together with a conventional modern web stack. The resulting system, Imaginify, supports object removal, background removal, recoloring, and generative fill with average processing times of 4-8 seconds and a 98% successful-request rate under light concurrent load, validating the architectural direction the survey proposed. Future work should extend load testing well beyond 10 concurrent users to characterize GPU-bound throughput limits at production scale; incorporate quantitative image-quality metrics such as FID or CLIP score alongside the current qualitative evaluation; explore model distillation or reduced-step sampling to bring processing time closer to real-time; and investigate attention-visualization or other explainability mechanisms to address the interpretability gap that remains open after this implementation.

ACKNOWLEDGMENT

The authors express their gratitude to their project guide, Prof. Varsha M. Gosavi, and to the Department of Computer Engineering, Smt. Kashibai Navale College of Engineering, Pune, for their consistent guidance, encouragement, and technical assistance throughout the completion of this work.

REFERENCES

1. J. Ho, A. Jain, and P. Abbeel, "Denoising Diffusion Probabilistic Models," in Advances in Neural Information Processing Systems (NeurIPS), 2020.
2. R. Rombach, A. Blattmann, D. Lorenz, P. Esser, and B. Ommer, "High-Resolution Image Synthesis with Latent Diffusion Models," in Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), 2022.
3. T. Brooks, A. Holynski, and A. A. Efros, "InstructPix2Pix: Learning to Follow Image Editing Instructions," in Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), 2023.
4. Hertz, R. Mokady, J. Tenenbaum, K. Aberman, Y. Pritch, and D. Cohen-Or, "Prompt-to-Prompt Image Editing with Cross Attention Control," arXiv preprint, 2022.
5. M. Sasirajan, S. Guhan, R. R. Mary, M. Maheswari, and S. R. Mary, "Image Generation with Stable Diffusion AI," International Journal of Computer Vision and Image Processing, 2023.
6. S. Kumar, T. Rode, Kunal, B. R. Ban, and M. P. Wankhade, "Stable Diffusion Using Deep Learning," International Journal of Emerging Technologies in Engineering Research, 2022.
7. X. Shuai, H. Ding, X. Ma, R. Tu, Y.-G. Jiang, and D. Tao, "A Survey of Multimodal-Guided Image Editing with Text-to-Image Diffusion Models," IEEE Transactions on Multimedia, 2023.
8. G. Al-Madhagy, T.-H. Hail, A. R. A. Alanzi, S. A. M. Yusof, and M. Alruwaili, "Software as a Service (SaaS) Cloud Computing: An Empirical Investigation," International Journal of Information Systems and Technology, 2021.
9. Y. Tang, W. Zhu, and Z. Gao, "Cloud-Based AI Model Deployment Using SaaS Architecture," International Journal of Cloud Applications and Computing, 2022.
10. N. Ruiz, Y. Li, V. Jampani, Y. Pritch, M. Rubinstein, and K. Aberman, "DreamBooth: Fine-Tuning Text-to-Image Diffusion Models for Subject-Driven Generation," in Proc. IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR), 2023.
11. Meng, Y. He, Y. Song, J. Song, J. Wu, J.-Y. Zhu, and S. Ermon, "SDEdit: Guided Image Synthesis and Editing with Stochastic Differential Equations," International Conference on Learning Representations (ICLR), 2022.
12. "Stable Diffusion-Based Image Editing as a Service: A Comprehensive Survey", IJEDR - INTERNATIONAL JOURNAL OF ENGINEERING DEVELOPMENT AND RESEARCH (www.IJEDR.org), ISSN:2321-9939, Vol.14, Issue 1, page no.88-90, February-2026, Available :https://rjwave.org/IJEDR/papers/IJEDR2601510.pdf